AD-A280 554

RR-94-21-ONR

DIAGNOSTIC ASSESSMENT OF TROUBLESHOOTING SKILL IN AN INTELLIGENT TUTORING SYSTEM

Drew H. Gitomer
Linda S. Steinberg
Robert J. Mislevy

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, NJ

April 1994

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1994

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Diagnostic Assessment of Troubleshooting Skill in an Intelligent Tutoring System

5. FUNDING NUMBERS
G. N00014-88-K-0304
PE. 61153N
PR. RR 04204
TA. RR 04204-01
WU. R&T 4421552

6. AUTHOR(S)
Drew H. Gitomer, Linda S. Steinberg and Robert J. Mislevy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-94-21-ONR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences
Code 1142CS
Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified/Unlimited

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT *(Maximum 200 words)*

This paper lays out the rationale and implementation of student modeling and updating in the HYDRIVE intelligent tutoring system (ITS) for aircraft hydraulic systems. An epistemic level of modeling concerns the plans and goals students are using to guide their problem-solving, as inferred from specific actions in specific contexts. These results update a student model constructed around more broadly defined aspects of system understanding, strategic knowledge, and procedural skills. Meant to support inferences that transcend particular problem states, this level of student modeling moderates feedback and instructional decisions in HYDRIVE. The applicability of this approach to student modeling in other learning domains is discussed.

14. SUBJECT TERMS
Bayesian inference networks, causal probability networks, cognitive models, HYDRIVE, intelligent tutoring systems.

15. NUMBER OF PAGES

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

Diagnostic Assessment of Troubleshooting Skill in an Intelligent Tutoring System[1]

Drew H. Gitomer, Linda S. Steinberg and Robert J. Mislevy
Educational Testing Service

Diagnostic Assessment of Troubleshooting Skill in an Intelligent Tutoring System

Drew H. Gitomer, Linda S. Steinberg and Robert J. Mislevy
Educational Testing Service

All intelligent tutoring systems (ITSs) are predicated on some form of student modeling to guide tutor behavior. Decisions based on inferences about what a student knows and does not know can affect the presentation and pacing of problems, quality of feedback and instruction, and determination of when a student has completed some set of tutorial objectives. In this paper, we describe a view of student modeling that, in the course of implementing principles of cognitive diagnosis, takes advantage of concepts and tools developed in the areas of probability-based reasoning, educational assessment, and psychometrics in an attempt to develop a generalizable framework for student modeling within intelligent tutoring systems.

Student models in an ITS can fulfill at least three functions. First, given a set of instructional options, a student model provides information suggesting which of the available choices is most appropriate for an individual (Ohlsson, 1987). ITS's, because they explicitly represent domains of knowledge and task performance, prescribe instruction that should be designed at a level of cognitive complexity that will lead to successful performance and understanding. Without explicit representation of task performance, instruction may be focused on non-essential features of the domain being tutored (e.g., Kieras, 1988). Second, student models in ITS's enable prediction of the actions a student will take based on an analysis of the characteristics of a particular problem state with respect to what the system infers about the student's understanding (Ohlsson, 1987). Given some inferred understanding of students and of problems, one ought to be able to more accurately predict future performance than if no model has been specified. The degree to which student actions conform to these predictions is an indication of the validity of the inferences made by the student model. Third, the student model enables the ITS to make claims about the competency of an individual with respect to various

problem-solving abilities. These claims are a shorthand that help to decide about whether a person is likely to be capable of negotiating a particular situation and can help the tutor make decisions about problem selection and exit criteria from a program of instruction.

In order to fulfill all three functions, we propose an ITS student model architecture that attempts to satisfy a set of cognitive and psychometric criteria that we believe to be essential to any successful student model, particularly those embedded in an ITS. These principles have become embodied in a system called HYDRIVE, an intelligent video-disc based tutoring/assessment system designed to facilitate the development of troubleshooting skills for the F-15 hydraulics systems[2].

Criteria for Student Modeling

The goal of the HYDRIVE student model is to diagnose the quality of specific troubleshooting actions and also to infer student understanding of general constucts such as knowledge of systems, strategies, and procedures that are associated with troubleshooting proficiency. In designing the student modeling component for HYDRIVE, we attempted to satisfy the following five criteria of student modeling.

1. Assessment of generalized constructs. Wenger (1987) describes three levels of information that can be addressed by an ITS. The *behavioral level* of information typically has been concerned with the correctness of student behaviors referenced against some model of expert performance. Early ITSs such as SOPHIE-I (Brown, Burton & Bell, 1975) contrasted student behaviors with domain performance simulations as a basis for offering corrective feedback. The *epistemic level* of information is concerned with particular knowledge states of individuals. Using techniques such as model tracing (e.g., Anderson, Corbett, Fincham, Hoffman, & Pelletier, 1992, Johnson & Soloway, 1985), and issue tracing (e.g., Lesgold, Eggan, Katz, & Rao, 1992), these tutors make inferences about the goals and plans students are using to guide their actions during problem solving. Feedback is responsive to "what the student is thinking." The *individual level* of information addresses broader assertions about the individual that transcend particular problem states.

Whereas the epistemic level of diagnosis might lead to the inference that "the student has a faulty plan for procedure X", the individual level of information might include the assertion that "the student is poor at planning in contexts A and B."

It is this individual level of information that has received the least attention in the field of intelligent tutoring assessment. Traditional psychometrics, on the other hand, has focused almost exclusively on claims about individuals while ignoring epistemic levels of information. An assertion is made, for example, that an individual has high ability in mathematics, yet the epistemic conditions that characterize high ability are never explicitly recognized. By recognizing and bridging between both individual and epistemic levels of information, an assessment model can have both the epistemic specificity to facilitate immediate feedback in a problem-solving situation, and also the generality of individual information to suggest the appropriate sequencing of problems, the moderation of instruction, and the determination of general levels of proficiency.

To meet this objective, the HYDRIVE student model is designed to make generalized claims about aspects of student troubleshooting proficiency based on detailed epistemic analysis of particular actions within the system. These generalized claims describe individual understanding at a level abstracted from any single problem solving situation. Abstractions, such as a student's *strategic understanding* , become the target constructs of the troubleshooting domain that are the focus of instruction.

2. <u>The student model as an implicit theory of performance</u>. ITS student models typically have been "runnable" in that they are designed to generate student performance and produce the same types of errors and successes that an actual student would if given a particular problem. The HYDRIVE student model's generalized abstractions are not runnable in the same sense. It will not generate specific actions, but it will predict the likelihood of occurrence for different classes and quality of actions. The student model is however, an implicit theory of

performance since the model-generated profile of student competencies predicts how students will perform on different problems and in different problem situations.

Such a theory of performance can also be viewed as a curricular goal structure. Lesgold (1988) has argued that ITSs, though they explicitly represent requisite knowledge to perform a task, have failed to articulate knowledge interrelationships in anything approximating a curriculum structure. The student model of HYDRIVE attempts to represent student understanding at the grain size of overarching curricular goals. Expert-like actions, for example, would lead to inferences that a student had good system understanding, an overarching curricular goal. The student model would not represent explicitly however, which specific system components and their features were well understood.

The HYDRIVE student model contains two levels of features. The first level can be construed as epistemic features, direct inferences of student understanding referenced to actions taken at a particular problem state. The second level of features represents the generalized constructs of individual proficiency. Links between the generalized constructs and the directly inferred features represent an implicit theory of performance in this domain. So, for example, the student model suggests that an individual with high strategic understanding is more likely to take an action that results in information about multiple components, when this is possible, than is an individual who is judged to have poor strategic understanding.

3. The student model as a predictor of actions. Typically, ITS student models have not supported prediction of actions based on higher-level assertions about individual competence. Prediction is more often confined to the relatively local level of plans, goals, and knowledge in highly specified contexts. An explicit goal of the HYDRIVE student model is to provide a mechanism for making predictions of student actions based on estimates of higher-order constructs. The ability to make such predictions creates the opportunity to directly test the adequacy of the model by evaluating how well student actions are predicted. The testability of student model

adequacy, particularly with respect to higher-order constructs, is a feature missing from most ITSs.

4. The student model as probabilistic. ITS modeling decisions have either been deterministic or at most, probabilistic in a limited sense. In deterministic models, a student is judged as having either evidenced or not evidenced some underlying skill or understanding via examining student behavior. For example, many of the bug-like approaches (Brown & Burton, 1978; Spohrer, Soloway & Pope, 1986), make definitive inferences that a student is operating under one conception or another.

Obviously, such inferences of unobservable reasoning processes can never be certain. To address uncertainty, a number of systems have adopted local probabilistic representation schemes that assign some likelihood values to inferences made by the student model. These systems do not use probabilistic reasoning to update inferences except at the most local levels. Updates follow relatively ad-hoc, albeit sensible updates of likelihood, that do not reflect the interdependencies of probabilities that should exist within a structural network that is governed by probability theory. Anderson's (Anderson & Reiser, 1985) LISP tutor is one such example of this approach.

Lesgold, Eggan, Katz & Rao (1992) have modeled student performance using a fuzzy variable methodology. Evaluated actions update unobservable variables in a consistent, but non-probabilistic manner. Though the rules of probability theory (e.g., $\Sigma p_1 ... p_n=1$) are preserved locally, probabilistic relationships between variables are not specified. This lack of specification precludes the testability of interdependencies among variables.

The HYDRIVE assessment scheme takes advantage of advances in probabilistic networks to characterize and assess the quality of a student model through the application of probability theory. Mislevy (Mislevy, 1993; Mislevy, Yamamoto, & Anacker, 1992) has presented the logic for the application of this methodology to issues of assessment. Essentially, it combines the statistical power of

probability theory to networks that are structures derived through the cognitive analysis of task domains. Probability theory provides a sound approach to evaluate, modify, and test student models predicated on cognitive understanding of task performance.

5. The student model as generalizable to other domains. The HYDRIVE model is designed to be generalizable to other domains aside from technical troubleshooting. If there exists a cognitive model of domain performance in which the interrelationship between features can be specified probabilistically, and if student behaviors within the tutor can be evaluated in terms of performance on some subset of those features, than this approach should be feasible. The power of this approach derives from the explicit representation of relationships between features, not from any particular qualities of the features themselves. Therefore, Mislevy has had success in modeling such tasks as arithmetic (Mislevy, 1993) and proportional reasoning (Beland & Mislevy, 1992), in addition to the current effort.

HYDRIVE's Design and Rationale

In this section, we overview the HYDRIVE system in order to introduce the context in which this student modeling approach was developed. HYDRIVE is designed to simulate many of the important cognitive and contextual features of troubleshooting on the flightline. Hydraulics systems are involved in the operation of flight controls, landing gear, the canopy, the jet fuel starter, and aerial refueling. Technicians in this career field diagnose and service F-15 problems on the flightline, where the aircraft takeoff and land. Their mission is to keep the aircraft flying as regularly as possible. In addressing problems, they typically isolate faulty components and replace them. Actual repair of any faulty component is performed by other individuals in a shop environment.

HYDRIVE presents problems as video sequences in which a pilot, who is about to take off or has just landed, describes some aircraft malfunction to the hydraulics technician (e.g., the rudders do not move during pre-flight checks). Once

the problem is presented, HYDRIVE's interface allows the student several options. The student can perform troubleshooting procedures by accessing video images of aircraft components and acting on those components. Alternatively, the student can choose to review technical support materials, including hierarchically organized schematic diagrams, which are available on line. Students can also make their own instructional selections at any time during troubleshooting, in addition to or in place of instruction that is recommended. A schematized version of the interface is presented in Figure 1.

Insert Figure 1 about here

The general structure of HYDRIVE is presented in Figure 2, with the modules responsible for student modeling highlighted. Students act on the aircraft through the interface. The state of the aircraft system, including changes brought about by user actions, is represented in the system model. The quality of student troubleshooting is monitored by evaluating how the student uses information in the system model to direct troubleshooting actions. As a result of decisions made by the student model, instructional help may be suggested by the tutor. The student model, then, is best understood in terms of its relationship to the system and instructional models.

Insert Figure 2 about here

The goal of creating an assessment scheme that represents an implicit model of student performance (Criterion 2) must rely on an understanding of the nature of task performance by individuals with different levels of expertise. Further, as an intelligent tutoring system, both the tutoring or instructional goals and the assessment constructs ought to derive from a common understanding. Therefore, the rationale for HYDRIVE's design was established through the application of the

PARI cognitive task analysis methodology developed in the Basic Job Skills Program of the Armstrong Laboratories (Means & Gott, 1988; Gitomer et al, 1992). The purpose of this analysis was to understand the critical cognitive attributes that differentiate proficient from less-proficient performers in the domain of troubleshooting aircraft hydraulic systems. PARI analysis is a structured protocol analysis scheme in which maintenance personnel are presented a problem and then asked to solve the problem mentally, detailing the reasons for their action (Precursor), and the Action that they would take. The technician is presented a hypothetical Result and then asked to make an Interpretation of the result in terms of how it modifies understanding of the problem. Technicians are also asked to represent their understanding of the specific aircraft system they are troubleshooting by drawing a block diagram of the suspect system.

Proficiency differences were apparent in three fundamental and interdependent areas: system understanding, strategic understanding, and procedural understanding, all of which are necessary for formation of an effective mental model of a system. These are the generalized constructs upon which the content of HYDRIVE is based. The coherence of the assessment approach, and the tutor itself, is due to the fact that the constructs monitored in the student model profile and the instructional goals all derive from the same PARI cognitive task analysis.

System understanding. System understanding consists of how-it-works knowledge about the components of the system, knowledge of component inputs and outputs, and understanding of system topology, all at a level of detail necessary to accomplish necessary tasks (Kieras, 1988). Novices did not evidence appropriate mental models, as represented by the block diagrams they were asked to draw, of any hydraulic system sufficient to direct troubleshooting behavior. In most cases, novices were unable to generate any mental model at all. The "models" they did generate generally included a small number of unconnected components that were so vague as to be of minimal use in troubleshooting. The operation of any given

aircraft system was essentially a black box for these technicians. Mental models for the experts, also represented by the block diagrams they were asked to draw, tended to be accurate representations of the specific aircraft system, including connections between components and between power systems. Experts' mental models generally evidenced a full understanding of how individual components operated within any given system, even though they did not understand the internal workings of these same components, which they had only to replace. Examples of expert and novice representations for the same problem (rudders fail to deflect with input) are presented in Figures 3 and 4.

Insert Figures 3 and 4 about here

Experts also demonstrated a principled sense of hydraulic system functioning independent of the specific F-15 aircraft. They seemed to understand classes of components beyond the specific instances found in a particular aircraft or aircraft system. Their knowledge was hierarchically organized according to the functional boundaries of the system. For a flight control system for example, hierarchical and generic clusters of components would include at least a switching system (for emergency backup), an electrically controlled input system, a hydraulic power source, and a set of hydraulic controls (the servo-actuators and related valves). At an even higher level, experts also understood the shared and discrete characteristics of flight control and other hydraulic-related aircraft systems.

The most important consequence of this type of understanding is that, in the absence of a completely pre-specified mental model of a system, experts are able to construct a mental model using schematic diagrams. They are able to flesh out the particulars given their basic functional understanding of how hydraulic systems work in the context of the aircraft. Experts are also able to use their knowledge of failure characteristics to help isolate a problem to a particular aircraft or power system. For example, intermittent failures have a higher likelihood of being

electrical rather than hydraulic in nature.

Strategic understanding. Novices did not employ very effective troubleshooting strategies either. That is, they demonstrated little ability for using system understanding to perform tasks that would allow them to draw inferences about the problem from the behavior of the system (Kieras, 1988). In many cases, the only strategy available to these individuals was to follow designated procedures in technical materials, even when it wasn't clear that the symptom matched the conditions described in the written manuals. While these materials, known as Fault Isolation Guides (FIs) can be useful tools, novices frequently fail to understand how an FI procedure serves to constrain the problem space. It is not always clear to the novice what information about the system is addressed by a particular FI procedure. Even in those cases where the technician evidences some system understanding, a serial elimination strategy, where components adjacent to each other are operated on in order, is frequently used. This strategy allows the technician to make claims only about a single component at a time. A space splitting strategy, conversely, dictates the use of actions that provide information about many components at one time, making this type of strategy much less costly. Novices do not evidence a strategic orientation that minimizes the costs of troubleshooting procedures while problem solving.

Expert strategies are much more effective, select approaches that maximize information gain and minimize the expense of obtaining such information. Experts try to use effective space-splitting strategies which isolate problems to a subsystem through the application of relatively few and inexpensive procedures that can rule out large sections of the problem area. They almost always attempt to eliminate and localize power system failures (eg., functional failure due to something like a blown fuse) first; then activate different parts of the system until they find the path along which the failure manifests itself; and finally localize the failure to a specific segment of this path (i.e., mechanical, electrical, hydraulic). The only exception to this general strategic model occurs when an exceptionally cheap action is available

that provides some information about the system. The ability to balance cost (measured in time to complete an action) and information benefit is one of the hallmarks of expertise in this domain. Experts are able to evaluate results in terms of their mental models of the system and make determinations of the integrity of different parts of the aircraft. When experts consult the FI guide, they do so as a reference to double check whether they may be overlooking a particular problem source. They may execute a recommended FI procedure, but never in a purely procedural and mechanical fashion. For experts, an FI action is immediately interpreted in terms of and integrated with their system mental model.

Those technicians with intermediate skills are quite variable in their use of strategies. When individuals have fairly good system understanding, they frequently evidence effective troubleshooting strategies. When system understanding is weak though, technicians often default to FI and serial elimination strategies. If intermediates have a basic understanding of troubleshooting strategy that is dependent on system understanding, then the implication for instruction for these individuals is to focus on system understanding. For novices, the evidence suggests that direct strategy instruction may also be necessary.

Procedural understanding. Every component can be acted upon through a variety of procedures which provide information about some subset of the aircraft. Information about some types of components can only be gained by removing and replacing (R&R) them. Others can be acted upon by inspecting inputs and outputs (electrical, mechanical, and/or hydraulic), and by changing states (e.g., switches on or off, increasing mechanical input, charging an accumulator). Some actions inherently provide information only about the component being acted upon, while other actions can provide information about larger pieces of the problem area, depending upon the current state of the system model. R&R procedures tend to provide information only about the component being operated upon.

As individuals gain expertise, they develop a repertoire of procedures that can be applied during troubleshooting. Novices are generally limited to R&R actions

and the procedures specified in the FI. They often fail to spontaneously use the information that can be provided from studying gauges and indicators and conventional test equipment procedures.

Experts are particularly adept at partially disabling aircraft systems and isolating major portions of the problem area as functional or problematic. For instance, rudders can be controlled through electrical and/or mechanical inputs. By disabling the electrical system, for example, a great deal of information about both the hydraulic and mechanical paths can be obtained.

The relationship between system, strategic, and procedural understanding. A mental model includes information not only about the inputs and outputs of components, but also available actions that can be performed on components. The tendency to engage in certain procedures or strategies is often a function of the structure and completeness of system understanding, rather than the understanding of strategies or procedures in the abstract. Failure to engage in space splitting may be attributable to one of several factors. First, the troubleshooter may not understand the system sufficiently to suggest appropriate points to split the system. Second, the individual may not have available appropriate actions (procedures) that will effectively divide the problem space. A third possibility is that the troubleshooter is simply unaware of how and when to use a space-splitting strategy. For those beyond the novice levels, the greatest reason for ineffective problem solving typically is attributable to poor system understanding. For the more novice individuals, there may even be an absence of a general aircraft system understanding that specifies the relationships between power systems.

Task analysis implications for assessment. This view of troubleshooting expertise has implications for student modeling and corresponding instruction in HYDRIVE. For assessment, failure to execute an effective troubleshooting action may, on the surface, appear to be a strategic failure. However, because a superficial strategic deficit may be due, in fact, to an impoverished system understanding, poor problem solving will contribute to a lower estimate of a student's system knowledge

as well as a lower estimate of strategic knowledge. If a student has exhibited strong strategic understanding on other problems for which good system understanding exists, then the likelihood is greater that the performance deficit on a new problem is directly attributable to a poor system knowledge. The student model must therefore represent the conceptual interdependencies that we assume to exist between different forms of understanding.

HYDRIVE's instruction focuses on effective system understanding and troubleshooting strategies rather than on optimizing actions to take at a given point in a problem. Ineffective actions raise doubts about a student's system understanding, which might suggest instruction targeted towards student construction of appropriate and useful system models. A key instructional strategy is to help students develop a hierarchical model of system understanding that is the critical feature of expert knowledge. HYDRIVE attempts to make this structure explicit through the use of hierarchical diagrams and organized verbal information. The claim is that effective troubleshooting strategies are more likely to be utilized in the presence of such a hierarchical structure.

Implementation of HYDRIVE's Student Model

There are three primary components to the HYDRIVE student model; the *action evaluator, the strategy interpreter* and the *student profile.* These three components depend on information from the system model to produce their results. The strategic goal of troubleshooting is to effectively reduce the problem area: to get as much information about components in the system model, so as either to eliminate them as sources of the failure or pinpoint the failure, in as efficient and cost-effective manner as possible. In HYDRIVE, students' actions are evaluated in terms of the potential information they yield given the current state of the system model. The action evaluator consults the current state of the system model and calculates the effects on the problem area of an action sequence performed by the student on the system model. The strategy interpreter makes rule-

based inferences about the student's apparent strategy usage based on the quality of information (i.e., quantity and type of problem area reduction) obtained from the action evaluator. Although obtained in a wide variety of situations that students arrive in as they work through a problem, these results are expressed in terms of a more abstract set of variables that are meaningful across situations. In the terminology of Mislevy (1993), these are the "observable variables" x. Not all elements of this vector need apply to all situations, but all updating of the student model variables will be mediated in their terms. The results of the strategy interpreter are then used to update the student profile, a network representation of student competence. The network element nodes and relationships are derived from the PARI analysis and are updated across actions and problems. In Mislevy's terms, these more abstractly-defined aspects of competence comprise the student model variables, β. As described below, a critical activity is specifying the probabilities that students having a given configuration of student-model values would take actions described as various possible values of relevant observable variables; that is, $p(x \mid \beta)$. Each of the student components is described below, but because action evaluation is based on information obtained from states and changes in the system model, we begin with a brief discussion of system modeling in HYDRIVE.

The system model. In HYDRIVE, the student uses the system model to simulate various aircraft states and explore the results of these simulations as a means of finding where in the system the problem resides. A system model is defined as a set of components that are connected by means of inputs and outputs. A component can have any number of inputs and outputs. Connections between components are expressed as pairs of components, the first being the component producing an output to the second in the pair which receives it as an input. These pairings are called edges and are also qualified by the type of power (electrical, hydraulic or mechanical) characterizing the connection. For example, the connection between a rudder and its actuator (the servomechanism which causes it

to move) would be left rudder servocylinder_left rudder (mechanical) because the actuator produces a mechanical output which the rudder processes as input. Every component has a small set of possible inputs. For example, the landing gear control handle can be in the *up* or *down* position. The output of a component is controlled by its input(s) and the internal state of the component. Given a set of inputs, the component will produce one or more outputs, the value of which depends on whether or not the component is working. For example, moving the landing gear handle to the *down* position will mechanically activate a relay which results in the creation of an electrical path that energizes the mechanisms associated with landing gear operation, assuming none of these components is failed. A failure may cause no output or an incorrect output to be produced.

Every component also has a set of actions (procedures) that can be performed on it. Some components can be set or manipulated (e.g., switches or control handles), others can be checked for electrical function (e.g., relays), and others can be inspected visually (e.g., mechanical linkages).

The system model processes the actions of the student and propagates sets of inputs and outputs throughout the system. A student activates the system model by providing input to the appropriate components and then has the option of examining the results of such actions by observing any other component of the system. Thus, a student can move the landing gear handle down and then go and observe the operation of the landing gear. If the landing gear does not move down, the student may decide to observe the operation of other components in order to begin to isolate the failure.

When a student uses the system model to simulate certain aircraft conditions and then observes the results of that simulation, information about the problem area (i.e., which components are still candidates as the source of the failure and which components have been eliminated as possibilities) is presumed available. If the pilot moves the control stick and the rudders move as the student might expect, then an inference can be drawn that all components involved in rudder operation

when controlled by the stick are functioning correctly and should be eliminated as sources of the problem. If, however, the rudders do not move as expected, then the student should be able to make the inference that some component is not working correctly along the path activated by the simulation between the control stick and the rudders. Observation of components at intermediate points along this active path can provide information about subsets of components involved in this particular way of operating the rudders. If an expected output is not produced at point x, then an inference can be made that the faulty component is somewhere between the point of control (e.g., the control stick), and the point of observation.

The action evaluator. For the hydraulics technician, the system model appears as an explorable, testable aircraft system in which a failure has occurred. All components belonging to this system are part of the initial problem area, represented as sets of input/output edges. When a student acts to supply power and input to the aircraft system, the effects of this input spread throughout the system model (as values propagated along a continuum of component edges), creating explicit states in a subset of components. This subset is called the active path. If one thinks of the system model as bounded on the one hand by the point(s) at which input is required to initiate system function (point of control), and on the other by its functionally terminal outputs, then an active path typically begins with the one and ends with the other, including all the connections in between. So, for example, an active path can be created for the steering system of an automobile by turning the steering wheel. This action creates an active path extending from the steering wheel (the input boundary, or a point of control of the system) to the tires (the output boundary of the system). For a power steering system the ignition switch is another point of control, since whether or not input is also supplied to turn the engine on affects the contents of the active path (one would be primarily hydromechanical, the other strictly mechanical).

The action evaluator considers every troubleshooting action from the student's point of view in terms of the information that can be inferred with respect

to effects on the problem area. The action evaluator, in updating its problem area, assumes that the student always makes the correct judgment about whether observations reveal normal or abnormal component states. If, for example, having supplied a set of inputs, a student observes the output of a certain component, which the system model 'knows' is normal, then the student is presumed to infer that all edges on the active path, up to and including the output edge, are functioning correctly and, therefore, remove them from the problem area. If the student, in fact, makes the correct judgment about the observation and the appropriate inferences from it concerning the problem area, then the dynamic problem area that the student model and the student hold correspond and troubleshooting continues smoothly. If, however, the student decides that the observed component output was unexpected, or abnormal, then, at least in the student's mind, all the edges in the active path remain in the problem area, any others would be elimininated, and the problem area maintained by the student model begins to diverge significantly from the one present in the student's mind. In this case, subsequent student actions and corresponding evaluations are likely to signal the need for instruction.

Figure 5 presents a grossly simplified hypothetical problem space for a hydraulics-like system. This system has two points of control which both send electrical signals to electrical components A and B respectively. Both of these signals are sent to an electromechanical component which outputs a mechanical signal to the mechanical component. Hydromechanical components A and B operate by receiving the mechanical signal as well as hydraulic power from hydraulic circuits A and B respectively.

Insert Figure 5 about here

In this hypothetical model, a number of active paths can be set up to isolate a fault. By activating point of control A, the entire system other than the path that

includes point of control B and electrical B are being tested. If the output from the hydromechanical components is unexpected, then the problem is clearly not associated with point of control B or electrical B edges. If expected output were to be obtained when point of control B is activated, then it is possible to infer that the locus of the fault is point of control A or electrical A, for other than these two component edges, the active paths overlap. Other discriminations can be made by selectively disabling hydraulics A and B and observing changes in the output of the hydromechanical devices. In HYDRIVE, the student can use a review function to help compare his or her dynamic idea of the problem area with that maintained by the student model.

The strategy interpreter. Actual strategy evaluation occurs by evaluating changes to the problem area, formally represented as **k**, the entire series of edges belonging to the system/subsystem where the problem occurs. As a student acts on the system model, **k** is reduced, with elements from **k** being removed as a result of an action sequence. If a failed component is on the active path, under the assumption that only one component fails at a time (a reasonable assumption in this domain), all edges other then those on the active path are eliminated from **k**. Upon inspection of any particular component on this path, the system model will also reveal a state which may or may not be expected from the student's perspective. The update of **k** stems from an inference that the fault has to be located within the active path and so all other components are removed from consideration. If, however, there is no failed component in the active path, then all edges in the active path are eliminated from **k**, while all other component edges remain in the problem area as candidate failure sources. The system model will return states that should be judged normal by the student for component edges along this active path. Also, an individual component is removed from **k** whenever the student selects a remove and replace action. Here, the assumption is that the replacement component is operational. However, with remove and replace, an inference can be made only about the output edges of the replaced component. No inferences are

possible for other components. The student's task is to reduce **k** until the problem is solved.

The method for reducing **k** is generalizable to any system that is comprised of components in which sequential flow of control can be defined. As long as one can make a judgment about the output state of a component, then inferences can be made about the state of components comprising a subset of the active path, from the point of control to the point of inspection.

When a sequence of actions results in new status information about more than one edge in the problem space, HYDRIVE designates the strategy as a type of **space-splitting**. HYDRIVE also differentiates between several forms of space splitting. There is **power system elimination**, which removes power system sources from the problem area (as in checking hydraulic pressure gauges or circuit breakers); there is **active path splitting**, which activates different combinations of components to achieve a particular system function (as in operating the rudders through the control stick and through the rudder pedals); and there is **power path splitting**, which either eliminates series of edges having the same power type or locates the failure to a particular power type (as in using electrical backup to replace mechanical function).

Other troubleshooting actions do not set up active paths and do not result in space splitting, but are discrete tests of single components. The most obvious is simply removing and replacing a component and observing whether the change results in a fix to the system. A **remove and replace** strategy is expensive both in terms of time and equipment, and is recommended only when there is a high degree of certainty that the replaced component is faulty. In the Figure 5 example, the electro-mechanical component could be replaced to test its functionality.

A **serial elimination** strategy refers to actions that only provide information about one edge at a time. A serial elimination strategy is inferred when one action provides information about one edge and the ensuing action provides information about an adjacent edge. Though the remove and replace strategy is a form of serial

elimination, HYDRIVE's designation is limited to actions that are not remove and replace actions (such as visual or electrical inspections).

An FI strategy is one in which the student follows procedures designated in an accessed FI guide for three consecutive actions. While such a strategy is not inherently problematic, it is clear that experts and novices use the FI in different ways. Therefore, the evaluation of a set of actions as an FI strategy will result in probes from the instructional model to ensure that the student understands the effects of actions taken.

Other evaluations do not actually infer strategies, but do make claims about the effectiveness of actions taken. **Redundant** actions are those that do not provide any new information about the problem. It should be noted that some actions are not costly to execute in terms of time or parts. In fact, experts often times will rerun a procedure to replicate and validate a finding. It is only when actions are costly and do not provide any new information that they are considered redundant. **Irrelevant** actions are those in which a student performs actions on components which are not at all part of any active path in the system of interest in the problem. Replacing the tires when an automobile won't start is an example of an irrelevant action.

The evaluation of the quality of a strategy is conditional upon the problem state at a particular point. While a remove and replace strategy is evaluated as poor when the problem state allows for space splitting, the same strategy is considered to be of better quality when the potential problem causes have been narrowed to one or two candidates. Therefore, within the strategy evaluator there exists a set of rules that characterize **k** in terms of the "best" strategy options that are available. Best strategies are strictly a function of the attributes of components in **k**, and are easily described. As an example, if components in **k** represent different power systems, then a potential strategy is to execute an action that will differentiate those components (a power space split). If all component edges in **k** represent one power system, such a strategy is not feasible.

HYDRIVE makes use of a strategic goal hierarchy to identify the optimal

strategy, given the current state of the problem area. Figure 6 contains HYDRIVE's strategic goal structure. The comparison of the student's strategy and the best strategy available, as calculated by the strategy interpreter, drives the instructional model which makes the strategic goal hierarchy embedded in the student model explicit to the student in the form of prompts, reminders and instructional exercises.

Insert Figure 6 about here

HYDRIVE employs a relatively small number of strategy interpretation rules (~25) to characterize each troubleshooting action in terms of both the student and the best strategy. An example of a student strategy rule is:

> *IF active path which includes failure has **not** been created and the student creates an active path which does **not** include failure and edges removed from **k** are of one power class, THEN the student strategy is power path splitting.*

An example of a best strategy rule is:

> *If **k** contains one or more hydraulic power systems, THEN the best strategy is power system elimination.*

<u>The student profile</u>. HYDRIVE uses the results of the strategy and action evaluator to update the student profile, represented as a network, using the ERGO (Noetic Systems, 1993) system. The student profile network that includes only a significant portion of the flight control system is presented in Figure 7. The nodes at the right are those that are directly updated through the strategy evaluation. These are thought of as observables. All other nodes can be thought of as constructs which have values determined, in terms of probability distributions for their possible values, by evidence captured by the observables. Once the observables are set by the strategy evaluation process, the remainder of the network is updated based on

probabilistic relations among nodes. There is an increasing level of abstraction and generality of inferences about students as one moves to the left of the figure.

The nodes and relationships in the network are derived from the PARI analysis. The PARI analysis supported the idea that proficiency could be characterized by knowledge of systems, strategies, and procedures, and that each of these broad areas could be characterized in terms of constituent parts. Analysis of individual differences in actions led to the association of constructs with particular observables. So, for example, the PARI data made it clear that an effective space-splitting action required knowledge of strategies, procedures, and the particular system being explored. The interdependencies evident in the PARI data are represented in the student profile network.

Insert Figure 7 about here

All of the nodes in the system, except for the direct strategy node (StratObs) are represented as having two states, each state having a probability associated with it. We are in the process of exploring more fine-grained distinctions among states. For example, Hawkes, Derry and Rundensteiner (1990), employing a fuzzy reasoning approach, have developed an ITS student model that makes use of seven levels of classification. For the observables, the states are *Positive* and *Negative*, for any strategy interpretation provides positive or negative evidence that some knowledge or skill is evident. When updated, they are assigned one of these two discrete states. The other nodes, those that are indirectly updated via the observables, are characterized by the states *Strong* and *Weak*, with a probability associated for each state.

The (splitable) node functions as a description of the current state of **k**, whether the remaining edges in **k** can be reduced by space splitting techniques or not. This is an important function, because the quality of an action can only be considered in the context of what is possible. Removing and replacing a

component, as already noted, is a costly procedure that provides limited information. Therefore, when space splitting is available, this type of action would be associated with less than expert troubleshooting. However, towards the end of a problem solution, when space splitting is no longer possible, remove and replace actions would be considered more positively.

The (StratObs) node takes on one of five values: *Space split, Serial elimination, Remove and Replace, Redundant and Irrelevant.* When the strategy evaluator makes an inference about the most recent sequence of troubleshooting actions, that inference is used to update each of the observables in a manner consistent with a conception of the interdependent nature of troubleshooting performance. As noted, a space splitting strategy not only indicates strategic understanding, but also indicates understanding of the system being troubleshot and the procedures used to effect the troubleshooting. Therefore, a number of observables will be updated positively when a space splitting strategy is inferred. On the other hand, a redundant action is negatively related with strategic understanding, system understanding and procedural skill. Corresponding observables would be assigned negative evidence in the case of a redundant evaluation.

The exact nature of the updating in any case is determined through probability-based inference; having specified the probabilities that a student with known competency values would take each of the potential actions in a given situation, then likeloods induced by the observation of a particular action are combined via Bayes Theorem with previous knowledge about the student to yield updated beliefs about the student-model variables. Thus, the same action can lead to qualitatively different updating when previous states of knowledge differ. For example, a redundant action taken when little is known about a student might lead to downgrading strategic understanding, system understanding, and procedural skill across the board. However, if we previously had evidence for good system understanding and procedural skill, but little evidence for strategic understanding,

the downgrading would appear mainly for the latter variable.

Once the observables are set, updating occurs as a function of the probabilistic relations specified in the network. Looking at the left side of Figure 7, *Proficiency* is a parent of *System Knowledge, Procedural Knowledge, and Strategic Knowledge.* The probability specification when the network is initially constructed is a response to the question "given that the student is proficient (strong), what is the probability that the student is strong in each of the respective knowledge areas" and also "given that the student is not proficient (weak), what is the likelihood that the student is strong in each of the respective knowledge areas?" If proficient people were always strong in system knowledge and non-proficient individuals were always weak in system knowledge, then the respective probabilities would be close to 0 and 1.

Such extreme values are seldom helpful in a network. First, it is rare that one can make such certain claims about anything based on someone's performance in an ITS. Second, the specification of such extremes in a network means that a single piece of evidence will have undue influence on the network. Any information that suggests someone has strong strategic knowledge would imply that the person is automatically proficient. By moderating the probabilities, one can temper the updating in the system so that multiple pieces of evidence influence any judgments.

The relative influence of a parent-child relationship is determined by the relative probabilities. Relationships having strong influence are characterized by child probabilities values that differ quite a bit for different parent conditions. Less influential relationships are characterized by child probability values that are more similar across different parent conditions. So, for example, because the PARI analysis showed that expert-novice differences were better described by strategic differences than by procedural differences (even novices have some expertise for different procedures), given a strong overall proficiency, the difference in probability values associated with strong and weak understanding , respectively is greater for strategic understanding than it is for procedural understanding. Those probability values are presented in Table 1. Increasing estimates of strategic understanding will

have a stronger impact on estimates of proficiency than will increased estimates of procedural understanding. Similarly, conditional probabilities of observable actions, given values of the student-model variables, were initially specified based on results from PARI traces. Having observed several acknowledged experts' and novices' solutions, we could begin to learn about the relative likelihoods that, say, an expert in a situation in which space-splitting was possible would in fact take a space-splitting action, compared to taking a redundant action, consulting the fault isolutation guide, and so on.

Insert Table 1 about here

Updating from instruction. While HYDRIVE's system model functions as a discovery world for system and procedural understanding, and its student model makes its evaluations based on an implicit strategic goal structure observed in expert troubleshooting, it is only in the instructional model that all of HYDRIVE's goals are made explicit. HYDRIVE's instructional model is driven by the comparison of the student strategy and what HYDRIVE 'thinks' is the best strategy under the prevailing conditions. The student is given great latitude in pursuing the problem solution; the instructional model intervenes with prompts or reminders (i.e., diagnostics) only when a student action constitutes an important violation of the rules associated with the strategic goal structure. As mentioned before, this is most likely to occur when the student's idea of the problem area and the student model's representation of same diverge in some dramatic way. Although HYDRIVE will diagnose and recommend some form of instruction, the actual presentation of any instruction is under direct control of the student who is free to take the instructional model's recommendation, choose other instruction, or continue troubleshooting without any instruction.

HYDRIVE's curriculum is directly informed by the cognitive attributes described in the student profile. The flow of control within the instructional model

is dictated by the assumption that the student must have adequate system knowledge (a 'runnable' model of the aircraft system) before selecting a troubleshooting strategy. Therefore, a student action which fails to reduce the problem area is first examined in the context of the student profile elements pertaining to system understanding. If these indicate a deficit, instruction is recommended to improve the student's mental model of the physical system. The results of many of these exercises (for example, the 'building' of an aircraft system/subsystem) provide direct evidence of the student's system understanding and cause the related profile elements to be updated. After the point that a student's profile elements indicate proficiency in system understanding, ineffective actions are considered in the context of strategic deficit and instruction shifts to emphasize and encourage HYDRIVE's strategic goal structure. Success or failure in certain of these exercises continues to update relevant profile elements.

Setting the probability values. In some situations where there is a large historical database, it is possible to determine empirically the conditional probabilities of observable variables given causal variables ("construct variables" in the present terms). In HYDRIVE, however, we do not have the luxury of analyzing large numbers of solutions from acknowledged experts and novices of various types. Initial values must be set subjectively, and revised as seen appropriate through model-checking activities. In essence, the objective is to encode a network structure and conditional probabilities specifications which correspond with experience to date not only locally (i.e., for a single given action-situation) but globabally (i.e., after accumulating evidence over a series of actions within a problem, then over a series of problems.) The HYDRIVE probabilities were set through an iterative process of making initial estimates, applying data obtained from the PARI analysis as proxies for what the student would do within the HYDRIVE tutor, and then evaluating the behavior of the network to determine whether all nodes were behaving sensibly in terms of the cognitive model. Initial probabilities were problematic in a number of ways. At times, student estimates would be updated too rapidly. At other times,

they wouldn't be updated despite actions that should have affected estimates of student competence. Other problems included updates moving in unexpected directions. Because all the probabilities are set at the individual node level, the behavior of the entire network is difficult to anticipate. However, by repeatedly applying data, and evaluating the network's behavior, probabilities can be tuned so that the system behaves in a manner consistent with human judgments of performance. These cycles of model building and model criticism are analogous to those required in the construction of, for example, medical expert systems (Andreassen, Woldbye, Falck, & Andersen, 1987)

Ultimately, as on-line data is obtained, the probabilities can be fine-tuned to an even greater degree. One of the values of this approach is that updates are propagated throughout the system, so that explicit predictions are made about the likelihood of a type of action occurring given a student profile. For example, a highly proficient student would be more likely to engage in space-splitting behavior given that space-splitting is possible than would a less proficient student. These likelihoods should be evident in the student profile and are able to be tested by evaluating student actions under these conditions. Discrepancies between predicted and observed actions will force refinement of the system.

Example student profiles. The updated profiles resulting from an ineffective and effective solution on a problem in the directional flight control system are presented in Figures 8 and 9, respectively. The ineffective solver first executed a number of actions that followed the FI guide. Following the FI does not result in any updating of the network, for following the FI is not inherently bad or good. Sometimes it makes sense and sometimes it doesn't. Simply using the FI to direct actions is insufficient to make a claim about the student. However, once the FI procedures failed to result in a solution, this solver immediately executed a number of remove and replace actions, a poor strategy at the outset of a problem. Following the remove and replace actions a number of serial eliminations were made. The solution was finally arrived at by removing and replacing the suspect component.

Insert Figures 8 and 9 about here

The expert solution began with a series of space splitting actions, followed by a number of serial elimination actions, some of which were taken when space splitting was no longer available. This person arrived at the solution in fewer steps than the less effective problem solver, concluding the problem by also removing and replacing the suspect component.

Differences in strategy usage and effectiveness of problem-solving are reflected in the networks in Figures 8 and 9. In reading the network, note that for all nodes except (StratObs), the upper bar is the probability of being strong on this node, and the bottom bar is the likelihood of being weak on the node. At the beginning of the problem, all likelihoods were at chance (.5).

As evidence accrues during problem solving some things to note in the network are:

1. The overall difference in likelihoods for the primary constructs of proficiency, strategic knowledge and system knowledge.

2. Differences in likelihoods for intermediate variables. For example, the effective solver is much higher on all of the strategic variables.

3. Relatively minor differences in the procedural likelihoods, an outcome of the probability structure that reflected the findings from the cognitive task analysis that experts and novices differed least in procedural skill.

4. Largest effects on variables in which the information is most direct, though likelihoods of related variables does change. For example, this problem was from the directional flight control system. Changes in estimates of strength were greatest for the directional system. Nevertheless, likelihoods for the lateral and longitudinal systems changed to a lesser extent, strengthening for the effective problem solver and weakening for the ineffective problem solver.

5. Changes in the expectations for the observables. Though it is difficult to see

in the figures, the StratObs distribution makes clear that there is a much greater expectation that the ineffective problem solver will take an action that is irrelevant or redundant than will the effective problem solver.

Controlling the model across problems. The preceding discussion has focused on updating a student model within a given problem, under the implicit assumption that a fixed state of competence is appropriate throughout the course of observation. All information about the student contributes equally to estimates of competence, regardless of when in the course of troubleshooting such information is obtained. The whole point of HYDRIVE, however, is to help students increase their competence! A mechanism to allow for change in the true status of student model variables is therefore necessary. To this end, we are adapting a recency strategy; that is, changes to the student-model variables effected by past problems will be fractionally reduced at the beginning of each problem, so that information from the current problem has more relative impact on our current beliefs than otherwise equally-informative information from past problems. Fractional reduction at the beginning of each problem implies a geometric rate of decay of information from past problems. To the extent that changes do occur over time, our current beliefs about student-model variables always lags their true status somewhat. This approach is more conservative and less risky than attempting to model learning explicitly, as in, for example, Anderson's LISP tutor (Anderson & Reiser, 1985).

Implications

We believe we have the beginnings of an assessment model that meets the five criteria set forth earlier in this paper. We are able to move from detailed analysis of discrete actions to make inferences about more general characteristics of an individual. This can be done because of an articulated cognitive framework of performance in this domain. The probabilistic features of this approach prevent ad hoc updating of variables and forces a clear specification of the relationship among

variables. The probabilistic network also allows for updating to work in two directions, parent-to-child and child-to-parent. The updating scheme allows for testing and evaluation of the student model, due to the explicit predictions that can be made. Most ITS student models are not capable of generating such predictions and are, therefore, incapable of being evaluated in the same way.

This type of student modeling appears to be generalizable to many other tutoring contexts. The most obvious transfer would be to other ITSs in troubleshooting domains. The rules of strategy evaluation are likely to be generalizable since their generalizability resides in the ability to explicitly define strategies in terms of an action's effect on **k**. While other domains may require the definition of strategies different from the one used by HYDRIVE, as long as these strategies can be referenced to changes in the state of **k**, or some similar representation, such generalization is quite straightforward.

ITSs more broadly, regardless of domain, typically have some form of strategy/action evaluator. What many are lacking is the bridge between an action evaluator and claims about the individual. However, it seems that links to the individual are necessary if we want to make generalizations from specific problem solving contexts to broader claims about competence and also if we want to direct instruction to issues that transcend particular problem states. Since assessment is fundamentally a process of making generalized inferences based on specific information, this type of approach may contribute to the development of assessment in the ITS world.

More generally though, this approach to assessment has implications for assessment in traditional pedagogical contexts. Features that support student modeling in HYDRIVE are critically important to, though too often absent from, successful classroom instruction. The first requirement is a clear and explicit representation of the domain, or structure of knowledge, to be learned. More than just isolated facts about a domain, the structure of knowledge is a representation of the interrelationships of concepts within a domain. Defining and addressing

explicit conceptual targets in classrooms is a significant challenge to educational reform in virtually all domains (e,g, Rutherford & Ahlgren, 1990; National Council of Teachers of Mathematics, 1989).

The second feature is a cognitive model of performance that permits inference of student understanding from task performance. The issue of how one makes valid judgments about student ability out of complex task performance is of central concern in the current educational and assessment debate (Messick, 1992). Part of the solution undoubtedly requires improvements in how evidence is collected and evaluated in classroom settings (e.g., Gitomer & Duschl, in press). Systematic and detailed exploration of student performance and its relationship to target features of domain understanding will be needed if a move towards problem-based learning environments is to succeed. It is worth noting that the difficulties in implementing the HYDRIVE assessment scheme were not particularly technical. By and large, the hurdles involved the explicit definition of the profile and the conceptual mastery of the relationship between student actions and the interpretations that could legitimately be generated based on those actions. These relationships were established through the cognitive task analysis that included a detailed understanding of the domain and performance within the domain. The quality of the cognitive task analysis is undoubtedly the most important feature of this, or any ITS assessment approach.

Mislevy and colleagues have developed prototype assessment models for characterizing proficiency in several relatively constrained domains. These efforts have included proportional reasoning (Béland & Mislevy, 1992; Mislevy, Yamamoto, & Anacker, 1992), signed number arithmetic (Thompson, & Mislevy, 1993), and mixed number subtraction (Mislevy, this volume). In each of these efforts, belief networks were created on the basis of cognitive analyses of task performance in the domain. Related efforts in physics problem solving are described by Martin & VanLehn (this volume).

It is important to recognize that this is not a recommendation that all

teaching of all domains pursue such a rule-based, systematic approach (Mislevy, in press). Certainly, this methodology is more appropriate for some disciplines than others. Equally certain, only a subset of any disciplinary focus would benefit from this type of approach. However, for those arenas of understanding that are highly structured, and that have clear rules for navigating within that structure, this form of curricular specification and assessment should prove to be beneficial.

References

Anderson, J. R., Corbett, A. T., Fincham, J. M., Hoffman, D., & Pelletier, R. (1992). General principles for an intelligent tutoring architecture. In J.W. Regian and V.J. Shute (Eds.), Cognitive approaches to automated instruction, (pp. 81-106). Hillsdale, NJ: Lawrence Erlbaum .

Anderson, J.R., & Reiser, B.J. (1985). The LISP tutor. *Byte, 10,* 159-175.

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S. K. (1987). MUNIN - A causal probabilistic network for interpretation of electromyographic findings. Proceedings of the 10th International Joint Conference on Artificial Intelligence, 366-372.

Béland, A., & Mislevy, R.J. (1992). Probability-based inference in a domain of proportional reasoning tasks. *ETS Research Report 92-15-ONR*. Princeton, NJ: Educational Testing Service.

Brown, J. S. & Burton, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.

Brown, J. S. Burton, R. R. & Bell, A. G. (1974). SOPHIE: a sophisticated instructional environment for teaching electronic troubleshooting. *BBN REPORT 2790.* Bolt Beranek and Newman, Inc. Cambridge, MA.

Gitomer, D. H., Cohen, W., Freire, L., Kaplan, R., Steinberg, L., & Trenholm, H. (1992). The Software Generalizability of HYDRIVE. Princeton, NJ: Educational Testing Service. (Armstrong Laboratories Progress Report).

Gitomer, D. H. & Duschl, R. A. (in press). Moving towards a portfolio culture in science education. S. Glynn & R. Duits (Eds.). Learning science in the schools: Research reforming practice. Washington, D.C.: American Association for the Advancement of Science.

Hawkes, L.W., Derry, S.J. & Rundensteiner, E.A. (1990). Individualized tutoring using an intelligent fuzzy temporal relational database. International Journal of Man-Machine Studies. 33, 409-429.

Johnson, W. L. & Soloway, E. (1985). PROUST: An automatic debugger for Pascal programs. Byte, 10, 179-190.

Kieras, D. E. (1988). What mental model should be taught: Choosing instructional content for complex engineered systems. In M. J. Psotka, L. D. Massey and S. A. Mutter (Eds.), Intelligent tutoring systems: Lessons learned, (pp. 85-111). Hillsdale, NJ: Lawrence Erlbaum.

Lesgold, A.M. (1988). Toward a theory of curriculum for use in designing intellgent instructional systems. In H. Mandl and A. Lesgold (Eds.) (pp. 114-137). Learning issues for intelligent tutoring systems. New York: Springer Verlag.

Lesgold, A. M., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. In J. W. Regian and V.J. Shute (Eds.), Cognitive approaches to automated instruction, (pp. 49-80). Hillsdale, NJ: Lawrence Erlbaum.

Means, B., & Gott, S. P. (1988). Cognitive task analysis as a basis for tutor development: Articulating abstract knowledge representations. In M. J. Psotka, L. D. Massey and S. A. Mutter (Eds.), Intelligent tutoring systems: Lessons learned, (pp. 35-58). Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. (ETS RR-92-39). Princeton, NJ: Educational Testing Service.

Mislevy, R.J. (in press). Test theory reconceived. Journal of Educational Measurement.

Mislevy, R.J. (1993). Probability-based inference in cognitive diagnosis. ETS Research Report 93-xx-ONR. Princeton, NJ: Educational Testing Service.

Mislevy, R.J., Yamamoto, K, & Anacker, S. (1992). Toward a test theory for assessing student understanding. In R.A. Lesh & S. Lamon (Eds.), Assessments of authentic performance in school mathematics (pp. 293-318). Washington, DC: American Association for the Advancement of Science. (Also available as RR-91-32-ONR. Princeton: Educational Testing Service.)

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: NCTM.

Noetic Systems (1993). ERGO v. 1.2 (software), Baltimore, MD.

Ohlsson, S. (1987). Some principles of intelligent tutoring. In R. W. Lawler and M. Yazdani (Eds.), Artificial intelligence and education (Volume 1), (pp. 203-237). Norwood, NJ: Ablex.

Rutherford, F. J. & Ahlgren, A. (1990). Science for all Americans. New York: Oxford University Press.

Spohrer, J.C., Soloway, E., & Pope, E. (1986). A goal/Plan analysis of buggy Pascal programs. Human-Computer Interaction, 1, 163-207.

Thompson, T. D. & Mislevy, R. J. (1993). An inference network representation for studying students' knowledge of signed number arithmetic. Presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

VanLehn, K. (1988). Student modeling. In M. C. Polson & J. J. Richardson (Eds.), Foundations of intelligent tutoring systems, (pp. 55-78). Hillsdale, NJ: Lawrence Erlbaum.

Wenger, E. (1987). Artificial intelligence and tutoring systems. Los Altos, CA: Morgan Kaufmann.

Notes

1. This work was originally presented at the Conference on Diagnostic Assessment, cosponsored by American College Testing and the Office of Naval Research in May 1993. We are grateful to Duan-Li Yan and Lauren Nuchow for their technical assistance in the development of the student profiles. We also thank Isaac Bejar for helpful comments on a previous version of the paper.

2. HYDRIVE has been generously supported by Armstrong Laboratories of the United States Air Force. We are indebted to Sherrie Gott and her staff for their contribution to this effort. The views expressed in this chapter are those of the authors and do not imply any official endorsement by any organizations funding this work.

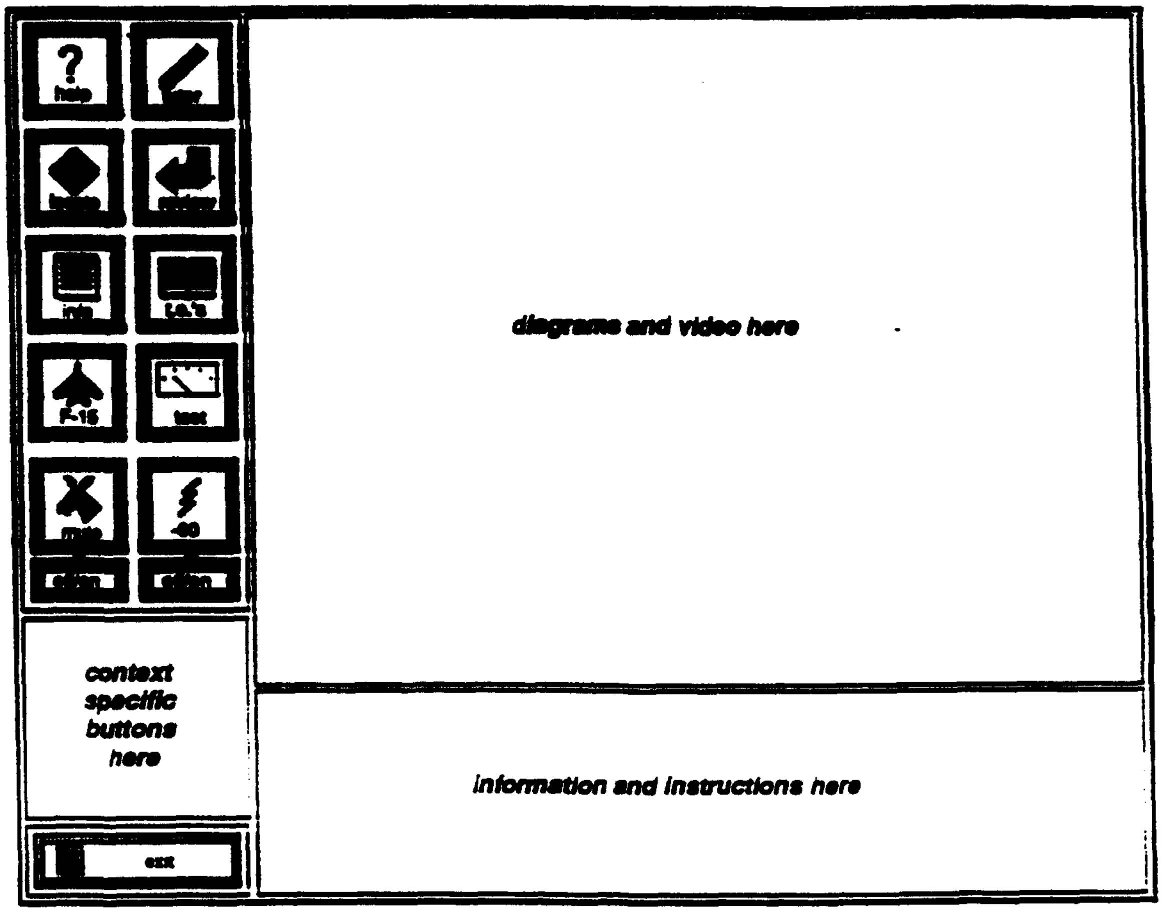
Figures

Table 1

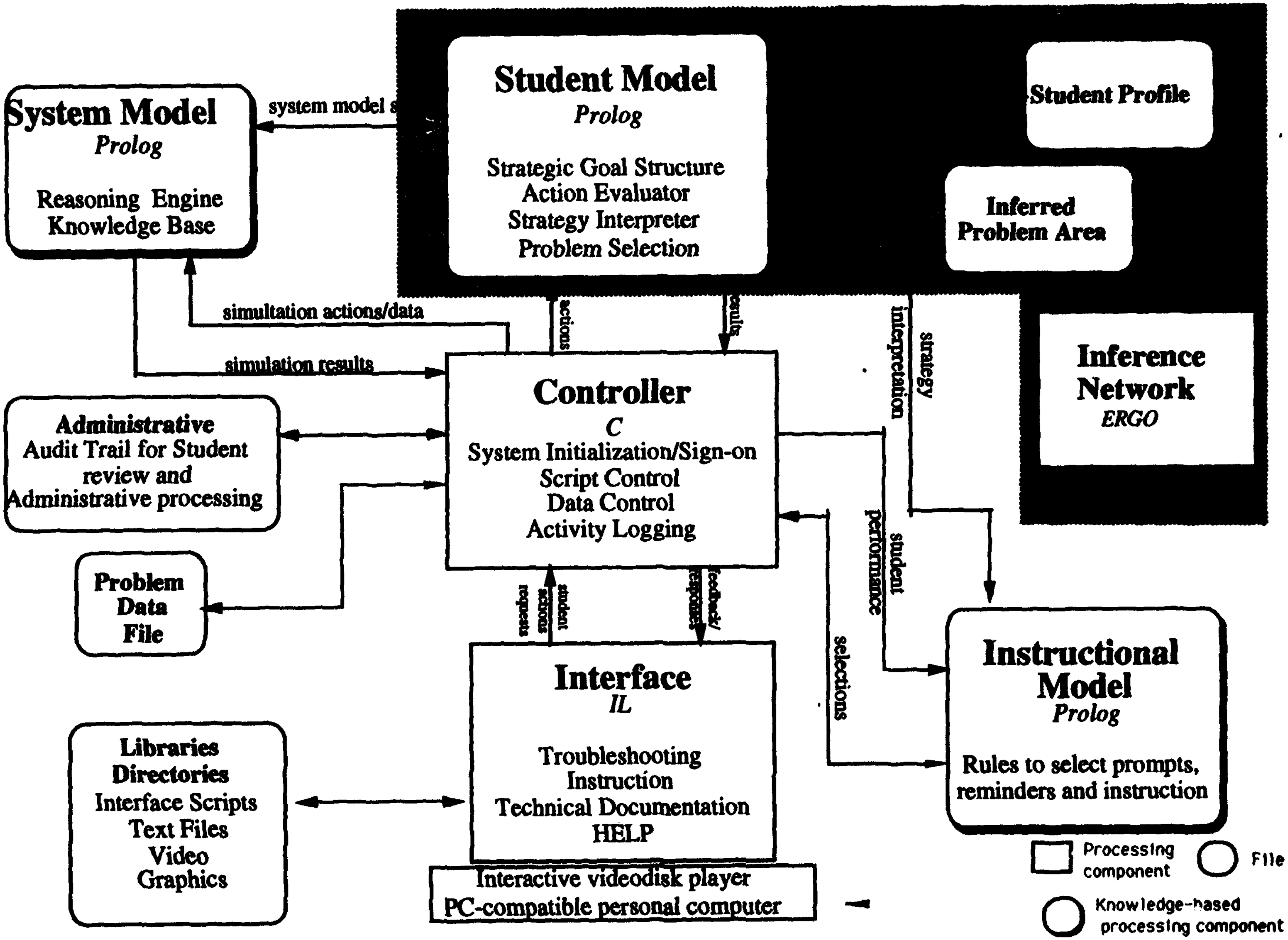
Network Probabilities for Strategic and Procedural Understanding Given Proficiency Level

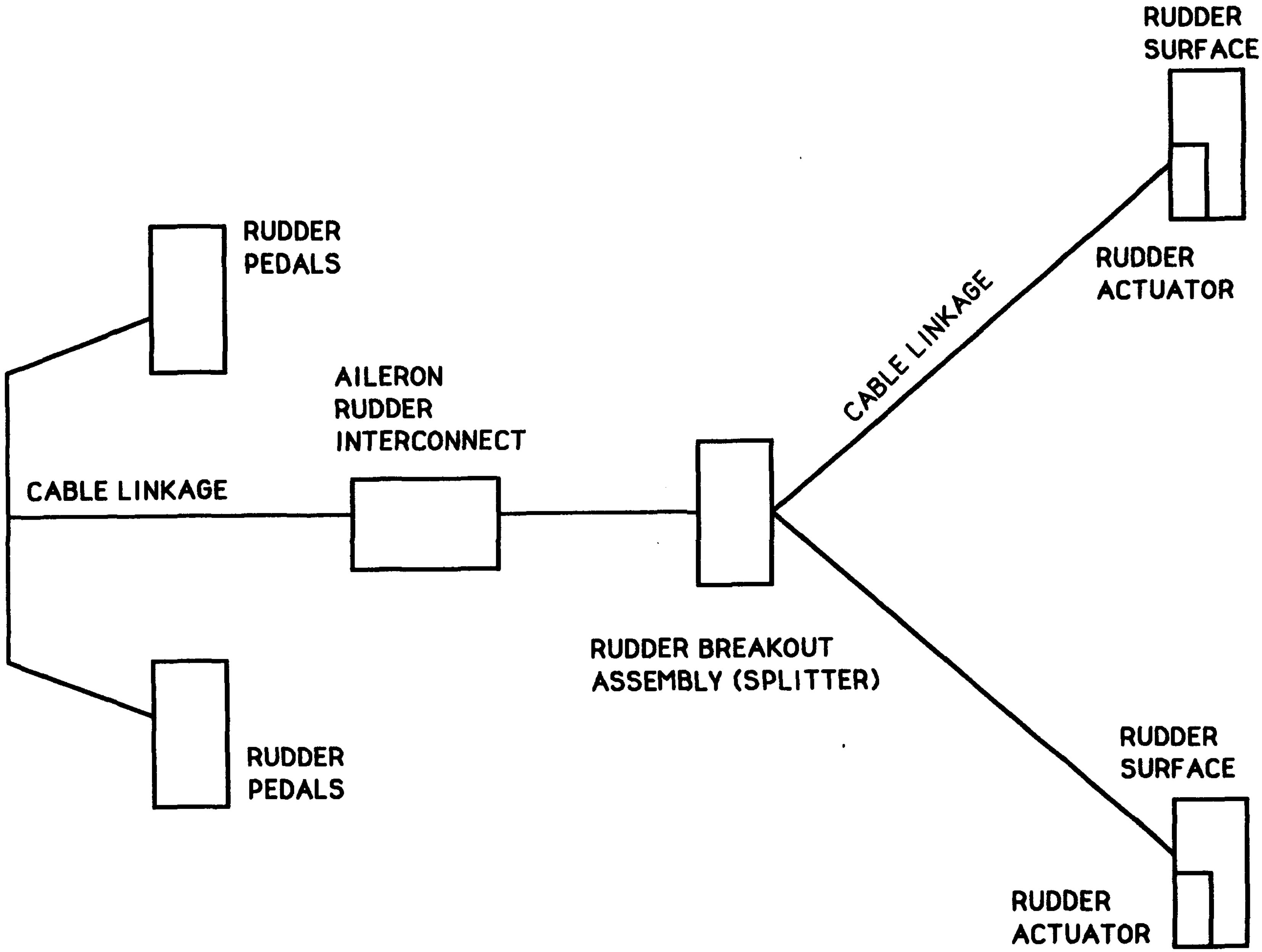
Child		Parent - Proficiency: Strong	Parent - Proficiency: Weak
Strategic Understanding	Strong	..80	..20
	Weak	..20	..80
Procedural Understanding	Strong	.52	.48
	Weak	.48	.52

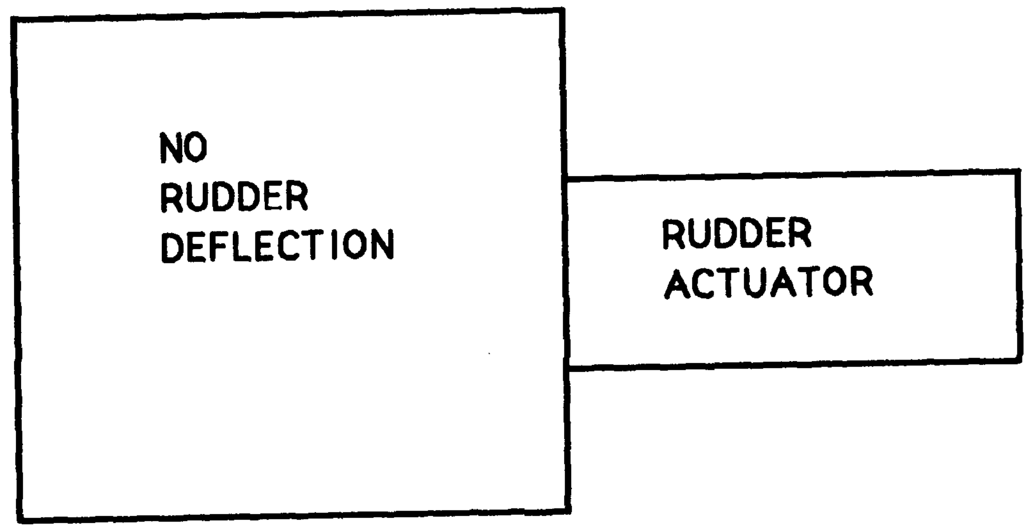
SYSTEM INTERFACE

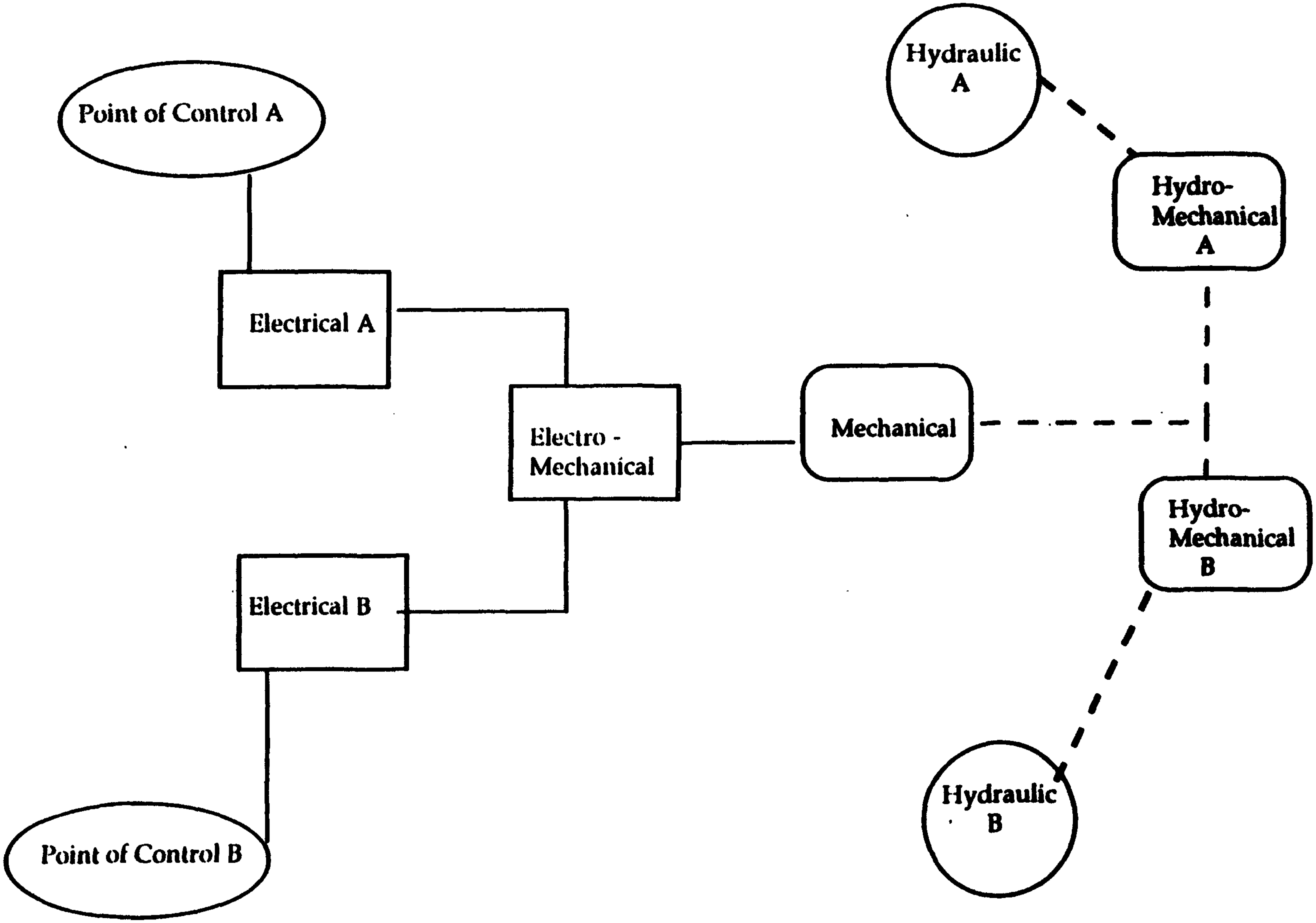


HYDRIVE Intelligent Tutoring System Diagram



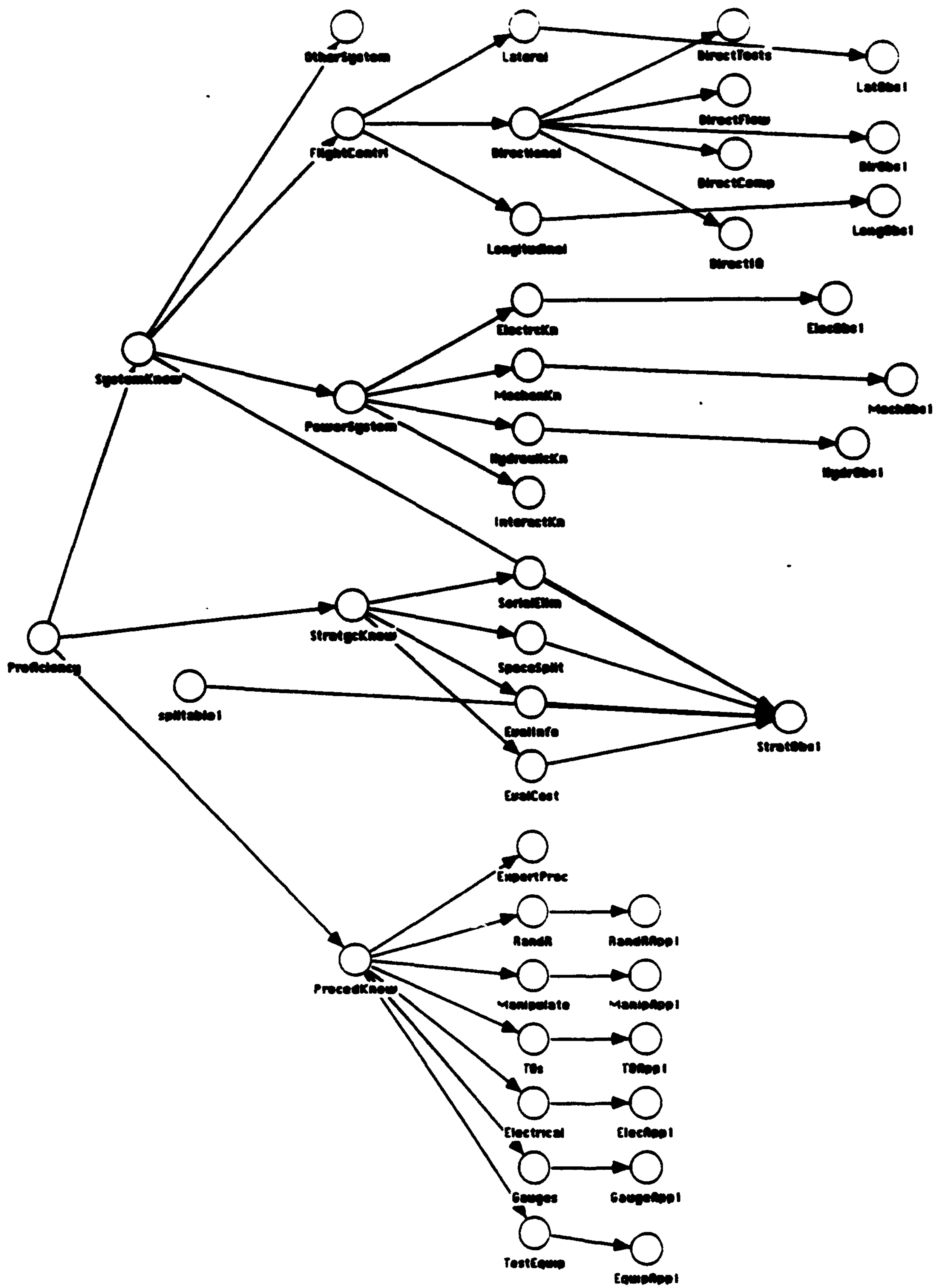

RUDDER
SURFACE
RUDDER
PEDALS
AILERON
RUDDER
INTERCONNECT
CABLE LINKAGE
CABLE LINKAGE
RUDDER
ACTUATOR
RUDDER BREAKOUT
ASSEMBLY (SPLITTER)
RUDDER
SURFACE
RUDDER
PEDALS
RUDDER
ACTUATOR

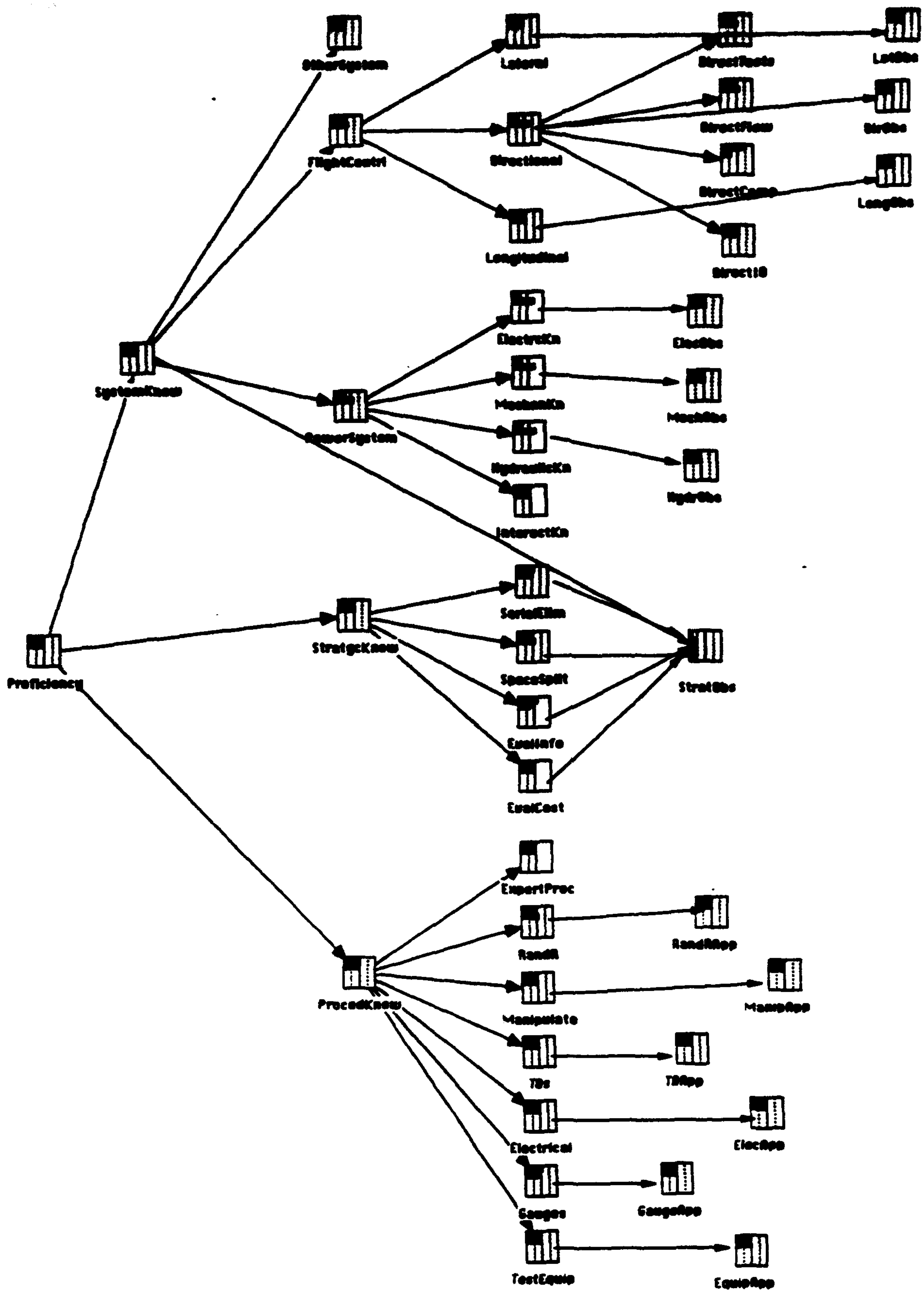

NO
RUDDER
DEFLECTION
RUDDER
ACTUATOR


Point of Control A
Electrical A
Electro - Mechanical
Mechanical
Hydraulic A
Hydro-Mechanical A
Hydro-Mechanical B
Hydraulic B
Electrical B
Point of Control B

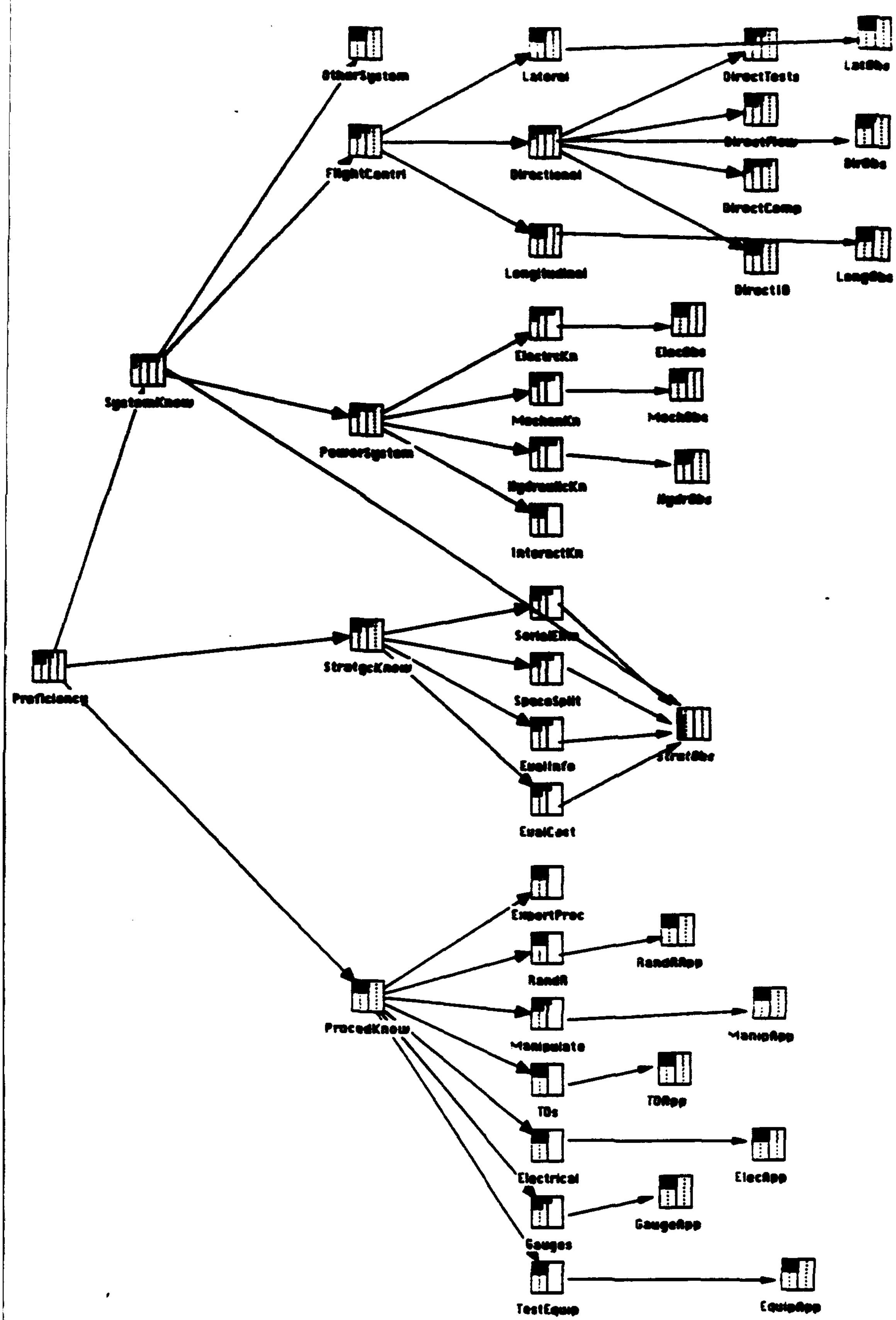
HYDRIVE's Strategic Goal Hierarchy

1. Power system elimination
2. Active path splitting
3. Power path splitting
4. Isolate failure within power path
    a. Serial elimination
    b. Remove and replace


OtherSystem
Lateral
DirectTests
LatObsl
FlightContrl
Directional
DirectFlow
DirObsl
DirectComp
Longitudinal
LongObsl
DirectIO
ElectricKn
ElecObsl
SystemKnow
MechanKn
MechObsl
PowerSystem
HydraulicKn
HydrObsl
InteractKn
SerialElim
StratgcKnow
SpaceSplit
Proficiency
splitabl
EvalInfo
StratObsl
EvalCost
ExpertProc
RandR
RandRAppl
ProcedKnow
Manipulate
ManipAppl
TOs
TOAppl
Electrical
ElecAppl
Gauges
GaugeAppl
TestEquip
EquipAppl



Update for ineffective problem solution



Update for effective problem solution

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Gregory Anrig
Educational Testing Service
Mail Stop 14-C
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr Isaac I Bejar
Educational Testing Service
Mail Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director
Life and Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
School of Education
Tel Aviv University
Ramat-Aviv 69978 ISRAEL

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Monterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Paul R Chatelier
PERCEPTRONICS
1911 North Ft Myer Drive
Suite 1100
Arlington VA 22209

Dr Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr William Crano
Department of Psychology
Texas A&M University
College Station TX 77843

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Meas Stat and Eval
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Dr Sharon Derry
Florida State University
Department of Psychology
Tallahassee FL 32306

Hei-Ki Dong
BELLCORE
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Educational Testing Service
Mail Stop 07-E
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd
Suite 550
Rockville MD 20850-3238

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Michael Kaplan
Office of Basic Research
US Army Research Institute
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick
Dept of Mathematics Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Logicon Inc (Attn: Library)
Tactical & Training Systems Div
PO Box 85158
San Diego CA 92138-5158

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr George B. Macready
Dept of Meas Stat & Eval
College of Education
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA 22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE 19716

Acad Prog & Research Branch
Naval Tech Training Command
Code N-62
NAS Memphis (75)
Millington TN 30854

Dr W Alan Nicewander
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Cntr for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC 20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660
(6 Copies)

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-01JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
NASA Ames Research Center
Mail Stop 239-1
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
American Council on Education
GED Testing Service Suite 20
One Dupont Circle NW
Washington DC 20036

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA 94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Dept of Industrial Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
WEST GERMANY

Dr David Wiley
Sch of Educ and Social Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Dept of Educational Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550